Information Technology Policy

Commonwealth Use of Public Generative Artificial Intelligence

Number
ITP-BUS014

Effective Date
September 21, 2023

Category
Business

Supersedes
None

Contact
RA-ITCentral@pa.gov

Scheduled Review
September 2024

1. Purpose

This Information Technology Policy (ITP) establishes basic guidelines for the use of Public Generative Artificial Intelligence (AI) by Commonwealth Employees and Contracted Resources (hereinafter referred to as "users").

2. Scope

This ITP applies to all offices, departments, boards, commissions and councils under the Governor's jurisdiction (hereinafter referred to as "agencies"). Agencies not under the Governor's jurisdiction are strongly encouraged to follow this ITP. Third-party vendors, licensors, contractors, or suppliers shall meet the policy requirements of this ITP that are applicable to the products and services provided to the Commonwealth.

3. Background

The field of AI has developed rapidly, and the availability and popularity of Generative AI has greatly increased. There is significant public interest in Generative AI because of its potential to empower creativity, innovation, and efficiency. However, the technology also presents risks and new challenges that the Commonwealth must navigate. This policy encourages users to engage with these technologies in a manner that is additive to their work and the services they provide to customers, residents, visitors, and industry, while addressing and mitigating risks inherent in its use. The Commonwealth will continually review policies and directives relating to the governance of Generative AI use in all its forms and evaluate Generative AI tools that may provide additional security while mitigating risks associated with this technology.

4. Definitions

Bias: Erroneous or prejudiced assumptions in artificial intelligence and machine learning processes that may affect generative output.

Generative Artificial Intelligence (Generative AI): Predictive algorithms that can be used to create new content including audio, code, images, text, simulations, and videos.

Private Generative AI: Generative AI tools that are specific to an entity or organization and their data. Private Generative AI tools can be developed in-house by an entity or organization for their own use or obtained from a third-party vendor. These systems are configured in a way that ensures an organization's data is segmented from other Training Data and accessible to only the entity or organization that owns it.

Public Generative AI: Generative AI tools that are openly available to multiple entities, organizations, or the general public and utilize widely sourced data from the internet as well as data from users or customers to train the Generative AI model. Public Generative AI tools do not guarantee the privacy of data input by users, entities, or organizations. Additionally, Training Data and models are not owned by a public organization unless otherwise noted.

Training Data: Data used to train a large language model and other predictive algorithms.

5. Objective

To promote the responsible use of Public Generative AI that allows the Commonwealth to benefit from the capabilities of these technologies while safeguarding Commonwealth IT resources and data and mitigating the associated risks.

6. Policy

If using Generative AI, users may only use a Commonwealth approved Public Generative AI tool. Public Generative AI tools that receive approval are available for use only in accordance with this policy. There are many Public Generative AI tools that offer different strengths and weaknesses. The Commonwealth will continue to evaluate and may approve additional Public Generative AI tools that may be of value to users.

6.1 Account Creation

Generative AI tools often require that users enter an email address to register and create an account. Users, who are utilizing an approved Public Generative AI tool for Commonwealth business purposes, shall use their Commonwealth e-mail address for registration and account creation purposes.

Once created, the account associated with a user's Commonwealth e-mail address shall be used solely for Commonwealth business purposes. Personal use of Public Generative AI from an account using a Commonwealth e-mail is prohibited.

Upon completion of the registration and the account creation process, users shall opt-out of data sharing and disable the chat history within the Public Generative AI system. If unable to opt-out, the user must contact the Office of Administration, Office of Information Technology (OA IT) prior to using the Public Generative AI system.

6.2 Risk Assessment

Generative AI is a versatile technology that can be used for a variety of purposes. Like any other tool, different use cases create different risks and rewards. While Generative AI tools are relatively new, many of the risks are the same as common internet or software-based tools. Examples of common risks with Generative AI tools are:

- Sharing private or confidential information in a Generative AI prompt.
- Generative AI outputs that are inaccurate or misleading in communications to the public or relying on inaccurate or misleading outputs to inform agency programs or policies.
- Reinforcing existing Bias in work products due to Bias in Generative AI outputs.
- Copyright infringement.

Because the use of Public Generative AI tools can pose significant risk depending on the information or data input into the tool, proper governance of such tools is required. When assessing whether to use an approved Public Generative AI tool, users should consider if the use case is high or low risk and high or low impact. High risk uses should be approached with additional review and governance and avoided when their impact is also minimal. Use cases that are low risk and high impact are potential opportunities to use Public Generative AI. Examples of such use cases are:

- **High risk/low impact (avoid):** Using Generative AI to draft an external facing communication that includes sensitive information for citizens and would have taken minimal time to write manually. Copying and pasting that output for use with minimal review.
- **Low risk/high impact:** Using Generative AI to compare a new and old version of a publicly available policy and asking the Generative AI tool to identify which sections have been modified, then, confirming the nature of these changes manually.

Generative AI & Coding: Generative AI tools can be an excellent coding resource with high impact. Public Generative AI tools used for coding purposes should be approached with caution, and special attention should be paid to risk assessment. Subject matter experience is required to properly validate Generative AI outputs, and this requirement is particularly necessary for coding use cases. Users must be cautious not to include production code or proprietary information in prompts, must assess vulnerabilities in code outputs, and must keep in mind that assessment of these outputs may require technical knowledge. Use of Generative AI in coding may result in more bugs or flaws in programs since it may gather the code from

flawed sources.

6.3 Accountability, Review, and Verification

When using any Generative AI for Commonwealth business purposes, **the user is accountable for any Generative AI outputs and must** review and verify all associated output content.

Qualifications to verify and review outputs: For a user to be able to review and verify outputs adequately, the user must have experience in the relevant topic area. For example, a software engineer may be able to verify the quality of code generated in a coding language in which the engineer specializes but may not be able to verify if a contract is legally sound without the requisite legal training.

Generative AI content should not be assumed to be accurate. At a minimum, users should review the output for:

- **Bias:** Since the data used to train Generative AI is vast, from a variety of sources, and not always vetted, outputs may contain inaccurate assumptions or stereotypes regarding certain individuals or communities.

- **Dated Information:** The data used to train Generative AI may have a fixed cutoff date, meaning any output generated will not reflect information available after a certain cutoff date.

- **Inaccurate Information:** Generative AI relies on Training Data. Training Data is vast and not always consistent or accurate. Inaccuracies in the Training Data may be included in the output generated by the Generative AI system.

  Inaccurate output can also be generated regardless of the Training Data. The Generative AI system may produce a confident response that appears plausible; however, the response is fabricated and divorced from reality (sometimes referred to as "hallucinations"). In one recent example, a user doing legal research using Generative AI was provided several court decisions, and the decisions provided by the Generative AI system turned out to be non-existent and completely fabricated.

- **Inappropriate Content:** If Training Data contains inappropriate content, the inappropriate content could appear in the Generative AI output.

- **Intellectual Property:** Generative AI tools continually ingest publicly available information for training purposes including information that may be subject to copyright. Copyrighted information could be inappropriately included in any output generated by the Generative AI system, creating intellectual property risks.

- **Confidential, Non-Public Information:** Since the data used to train Generative AI is vast, from a variety of sources, and not always vetted, outputs may contain confidential, non-public information.

6.4 Disclosure

Users shall be transparent about their use of Generative AI and must disclose to customers, residents, visitors, and industry when Generative AI has been used to

generate content that may be public facing or shared externally. Generative AI use must be disclosed even if it was only used to generate a portion of the content. The disclosure shall be prominently displayed and include an indication that the content was generated either entirely or in part by Generative AI and identify the Generative AI system and version that was used.

Example: "ChatGPT-3.5 was used in the creation of this document."

6.5 Data Privacy

Users should not have any expectation of privacy when interacting with Public Generative AI tools. Any data or information that users would not include in public facing documents or emails should never be entered into any Public Generative AI tools.

6.6 Prohibited Uses

6.6.1 Production Code and Proprietary Information

While it is acceptable to use Public Generative AI tools to modify or interpret code or other content, **under no circumstances should production code or proprietary information be used in prompts.**

6.6.2 Non-Text Outputs

Users shall not utilize Public Generative AI for non-text-based outputs. Most Generative AI platforms can generate images, video, audio, or other types of content. However, the risks related to inadvertently including other's intellectual property or generating offensive content are significantly higher and more difficult to detect than with text-based outputs. Structured data, numbers, code, and different languages are acceptable outputs from Generative AI only so long as the output is properly reviewed and verified by users with the appropriate expertise.

6.6.3 Private and Sensitive Data

No class C data, as defined in *ITP-INF015, Policy and Procedures for Identifying, Classifying, and Categorizing Commonwealth Electronic Data*, may be input into any Public Generative AI prompt, tool, or system. This includes, but is not limited to:

- Sensitive Security Information
- Personal Identifiable Information (PII)
- Protected Health Information (PHI)
- Regulated Data – Such as data from or regulated by:
  - Social Security Administration (SSA)
  - Internal Revenue Service (IRS)
  - Centers for Medicare & Medicaid Services (CMS)
  - Criminal Justice Information (CJI)
  - Criminal History Record Information Act (CHRIA)
  - Family Educational Rights and Privacy Act (FERPA)
  - Payment Card Industry Data Security Standard (PCI DSS)
- Confidential or Non-Public Information
- Privileged Information

- Prerequisite-Required Information

Additionally, any non-public records or information that would be considered privileged or exempt from access under the *Right-to-Know Law (RTKL), 65 P.S. §§ 67.101, et seq.* may not be input into any Public Generative AI prompt, tool, or system.

6.6.4 Decision Making

Generative AI outputs are not to be used to make decisions for or on behalf of employees. Employees may use Generative AI outputs to inform a larger decision-making process, but ultimately the Commonwealth employee or official remains the final decision maker. Users must review and verify all output produced with the assistance of Generative AI. The user will be accountable for any decision-making based upon such output. Generative AI cannot make reliable subjective or value-based judgments and may not be used for such purposes.

- For example, do not use generative AI to make final decisions that affect employment.

6.7 Acceptable Uses

Examples of acceptable uses of Public Generative AI include:

- Drafting a job posting or job description
- Summarizing or paraphrasing a writing
- Taking a technical answer to a question and rewriting it in customer-friendly language
- Creating an outline for a memo or other communication
- Brainstorming icebreakers for a meeting

The examples provided above assume that the Generative AI tool has been approved for use; only public, non-confidential data is involved; and proper review and verification is completed as outlined in section 6.3 of this policy.

This is not a comprehensive list of the permitted uses, but rather illustrates some common lower risk use cases.

7. Responsibilities

7.1 Agencies shall:

Comply with the requirements as outlined in this ITP.

7.2 Office of Administration, Office for Information Technology shall:

Comply with the requirements as outlined in this ITP.

7.3 Third-party vendors, licensors, contractors, or suppliers shall:

Comply with the requirements as outlined in this ITP that are applicable to the products or services they are providing to the Commonwealth. If the products or services being provided by the third-party vendor, licensor, contractor, or supplier do not fall within the scope of this ITP, compliance is implied. If the third-party

vendor, licensor, contractor, or supplier subsequently deploys products or services that fall within the scope of this ITP in the future, compliance with the policy is required.

8. Related ITPs/Other References

- Definitions of associated terms of this policy are published on the Office of Administration's public portal: http://www.oa.pa.gov/Policies/Pages/Glossary.aspx
- Commonwealth policies, including Executive Orders, Management Directives, and IT Policies are published on the Office of Administration's public portal: http://www.oa.pa.gov/Policies/Pages/default.aspx
- *Management Directive 205.34 Amended, Commonwealth of Pennsylvania Information Technology Acceptable Use Policy*
- *ITP-BUS012, Artificial Intelligence General Policy*
- *ITP-INF015, Policy and Procedures for Identifying, Classifying, and Categorizing Commonwealth Electronic Data*
- *ITP-SFT001, Software Licensing*
- *Right-to-Know Law* (RTKL), *65 P.S. §§ 67.101, et seq.*

9. Authority

Executive Order 2016-06, Enterprise Information Technology Governance

Executive Order 2023-19, Expanding and Governing the Use of Generative Artificial Intelligence Technologies Within the Commonwealth of Pennsylvania

10. Publication Version Control

It is the users' responsibility to ensure they have the latest version of this publication, which appears on https://itcentral.pa.gov for Commonwealth personnel and on the Office of Administration public portal: http://www.oa.pa.gov/Policies/Pages/default.aspx. Questions regarding this publication are to be directed to RA-ITCentral@pa.gov.

11. Exemption from this Policy

In the event an agency chooses to seek an exemption from the guidance within this ITP, a request for a policy waiver shall be submitted via the enterprise IT policy waiver process. Refer to ITP-BUS004 *IT Policy Waiver Review Process* for guidance.

This chart contains a history of this publication's revisions. Redline documents detail the revisions and are available to CWOPA users only.

Version	Date	Purpose of Revision	Redline Link
Original	09/21/2023	Base Document	N/A